Measuring Economic Growth from Outer Space: A Comment

Berliant, Marcus and Weiss, Adam

Washington Univeristy in St. Louis, The Urban Institute

1 June 2013

Measuring Economic Growth from Outer Space: A Comment

Marcus Berliant* and Adam Weiss†

June 2013

Preliminary Draft
Comments Welcome

Abstract

We examine spatial econometric issues arising from the model specification in Henderson, Storeygard and Weil (2012), that uses night light data to proxy for missing or unreliable GDP growth data.

JEL numbers: C21, C23 Keywords: GDP, Night light data, Spatial autocorrelation

*Department of Economics, Washington University, Campus Box 1208, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. Phone: (314) 935-8486, Fax: (314) 935-4156, e-mail: berliant@artsci.wustl.edu

†The Urban Institute. e-mail: weiss.l.adam@gmail.com

1 Introduction

Henderson, Storeygard and Weil (2012) employ night light data to augment more standard measures of GDP growth, particularly for countries where data is unreliable or missing, or for subnational units. Here we examine some spatial econometric issues that arise with this approach. Our comments arise both from the purely empirical point of view as well as from elementary demand and trade theory. We begin with the theory in Section 2, turning later to the empirics in Section 3. We end with the implications in Section 4.

2 Analysis

To begin, we reiterate the relevant notation from the paper that is our focus. Let y be the growth in true, real GDP, and let x be the growth in night lights. Let z denote measured growth in GDP. The subscript j denotes a country. Equation (1) of Henderson, Storeygard and Weil (2012) gives the specification of measurement error in the data:

$$z_j = y_j + \varepsilon_{z,j} \tag{1}$$

where $\varepsilon_{z,j}$ is measurement error. Equation (2) of Henderson, Storeygard and Weil (2012) is the statement of a basic relationship:

$$x_j = \beta y_j + \varepsilon_{x,j} \tag{2}$$

Here, $\varepsilon_{x,j}$ is the error term, whereas β is the elasticity of (growth in) lights with respect to (growth in) income. This equation is interpreted as a purely statistical relationship.

In contrast with that paper, we interpret this as a statement of a derivative of the (log) demand relationship for light. Suppose that X is light consumption, Y is income, and P is the price of electricity. Suppose that demand takes the functional form:

$$\begin{aligned} X &= Y^{\beta} \cdot P^{\alpha} \\ \beta &> 0, \alpha < 0 \end{aligned}$$

Taking logarithms of both sides and then the derivative with respect to time, and denoting growth rates by lower case variables, we obtain:

$$x = \beta y + \alpha p$$

There is an omitted variable relative to equation (2), namely the percent change in the price of electricity, the same as the percent change in the price of light. Then we would like to write

$$\begin{aligned} x_j &= \beta y_j + \alpha p_j + \varepsilon'_{x,j} \\ &= \beta y_j + \varepsilon_{x,j} \\ \text{where } \varepsilon_{x,j} &= \alpha p_j + \varepsilon'_{x,j} \end{aligned}$$

Since z denotes the measured growth of GDP, y is replaced with z from equation (1):

$$x_j = \beta z_j + \alpha p_j - \beta \varepsilon_{z,j} + \varepsilon'_{x,j}$$

Equation (3) is the basic relationship estimated in Henderson, Storeygard and Weil (2012, Table 2):

$$z_j = \widehat{\psi} x_j + e_j \tag{3}$$

In terms of our notation, and inserting a time index for clarity (as panel data is used), we obtain the following expression:

$$z_{jt} = \frac{1}{\beta} x_{jt} - \frac{\alpha}{\beta} p_{jt} + \varepsilon_{z,jt} - \frac{1}{\beta} \varepsilon'_{x,jt}$$

In other words,

$$e_{jt} = -\frac{\alpha}{\beta} p_{jt} + \varepsilon_{z,jt} - \frac{1}{\beta} \varepsilon'_{x,jt}$$

We wish to raise two spatial econometric issues with this regression.

First, although the growth in the price of electricity is not observed, such growth might be correlated between, for example, neighboring countries i and j. If p_{it} and p_{jt} are correlated, then there is a classical spatial autocorrelation/omitted variable problem. This results, for example[1], in a biased estimate of the key parameter $\widehat{\psi}$ *provided that* x_j and p_j are correlated.[2] Of course, given a demand equation, it is quite natural that growth in use of lights be (negatively) correlated with growth in the price of electricity. Since we don't have price data to insert into the regressions, if growth in light use is negatively correlated with growth in price of electricity in a location, and growth in electricity price in one location is positively correlated with the growth in electricity price in neighboring locations, then we can employ spatially lagged light use to proxy for these correlations, particularly for the omitted variable: growth in the price of electricity.

[1] Further implications will be discussed in the final section.

[2] As pointed out in Henderson, Storeygard and Weil (2012), there is another, independent reason $\widehat{\psi}$ might be biased. We shall return to this in Section 4.

Second, it seems clear to us that due to trade between countries close in distance, growth in real GDP at a given time could be correlated across space. In particular, higher income in one country can lead to higher demand for an adjacent country's products, thus raising income in the adjacent country. In other words, z_{it} and z_{jt} are correlated. This can also lead to biased estimates due to the omission of spatially lagged (and weighted) endogenous left hand side variables from the right hand side of the regression. Evidently, this issue does not arise in regressions where lights appear on the left hand side of the regression, but it does arise when growth in GDP is put on the left hand side whereas night lights are moved to the right hand side. This second issue is covered by inserting spatially lagged dependent variables on the right hand side of the regression.

Formally, the preceding two paragraphs amount to:

$$z_j = \widehat{\psi} x_j + e'_j$$

where

$$e'_{jt} = \lambda z_{it} + \frac{\gamma}{\beta} x_{it} + \varepsilon_{z,jt} - \frac{1}{\beta} \varepsilon'_{x,jt}$$

and countries i and j are neighbors.

Our goal is to address these issues, beginning with the basic regressions run in the paper, to see what affect this has on the use of light data to predict GDP. Our first focus is on reconstructing column 1 of Table 2 in Henderson, Storeygard and Weil (2012). Then we shall draw the implications for the analysis.

3 Estimation

Formally, what we have is a Spatial Durbin Model (SDM), that incorporates both the omitted price variable indirectly using the spatially lagged independent variable night lights as a proxy, and the spatially lagged dependent variable on the right hand side. Recalling that panel data is used, formally we write the econometric model as:

$$Z_t = \lambda W Z_t + X_t \psi + W X_t \frac{\gamma}{\beta} + u_t$$

where W is a spatial weight matrix, Z_t and X_t are the vectors of cross sections of the respective variables at time t, and $u_t (= \varepsilon_{z,jt} - \frac{1}{\beta} \varepsilon'_{x,jt})$ is the error vector.

An important special case of SDM is when $\gamma = 0$, namely there is spatial autocorrelation only in the dependent variable. This special case is called the Spatial Autoregressive Model (SAR):

$$Z_t = \lambda W Z_t + X_t \psi + u_t$$

Tables 1 and 2 contain our empirical findings.[3] The first column replicates the basic regression of Henderson, Storeygard and Weil (2012), column 1 of their Table 2. Unfortunately, we are unable to use their entire sample, as data is missing for some countries in some time periods, rendering the spatial weighting matrix that we must use to test for spatial econometric purposes difficult. Thus, we censor the countries for which data is incomplete, resulting in a smaller cross section sample size of 150. So for comparison purposes, in column 2 of our Table 1 we perform the same regression as in column 1 but for the smaller sample size. In Table 2, we report spatial test statistics for the appropriate regressions in Table 1. Throughout our application, the spatial weighting matrix that we use is simply a contiguity matrix with elements 0 and 1, where 1 is used to denote a geographic neighbor and 0 is used to denote the complement.[4] For the remaining regressions/columns, a spatial weighting matrix is necessary. In column 3 of Table 1, we run the GMM version of column 2 that also generates test statistics for model specification. We find strong evidence of positive spatial autocorrelation in the error terms of this regression, as seen (for example) in Moran's I statistic. In column 4, we run the GMM correction for the misspecification, incorporating the possibility of spatial autocorrelation in the fixed effects. In column 5, we run a SAR with fixed effects, whereas in column 6, we run the SDM model with fixed effects.

[3] The Stata code and spatial weight matrix used to generate the tables are available from the authors upon request.

[4] It is possible that the use of geographic distance in place of this crude measure might yield different results.

Table 1

Dependent Variable - Growth in Real GDP

	ln(GDP)	ln(GDP)	ln(GDP)	ln(GDP)	ln(GDP)	ln(GDP)
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS restricted sample	GMM initial unweighted	GMM[5] fully weighted	SAR	SDM
ln(lights/area)	0.277	0.267	0.267	0.267	0.301	0.369
Standard error	[0.031]	[0.031]	[0.011]	[0.011]	[0.033]	[0.041]
Observations	$3,015$	$2,550$	$2,550$	$2,550$	$2,550$	$2,550$
Countries	188	150	150	150	150	150
R^2	0.769	0.772	0.9994	0.9994	0.9990	0.9983

All regressions include a constant and space and time fixed effects. Standard errors are robust, clustered by country.

Table 2

Spatial Test Statistics

(p-values in parentheses)

Statistic	(3)	(4)	(5)	(6)
	GMM initial	GMM weighted	SAR	SDM
Moran's I	0.2235 (0.00)	0.2236 (0.00)	-0.0804 (0.00)	0.7127 (0.00)
Geary	0.9054 (0.0303)	0.9055 (0.0306)	0.9946 (0.8785)	0.3561 (0.00)
Getis-Ord	-0.6392 (0.00)	-0.6394 (0.00)	0.2298 (0.00)	-2.0384 (0.00)
LM Lag (Anselin)	133.94 (0.00)	134.62 (0.00)	0.0000 (1.00)	0.1026 (0.7488)
Akaike IC	0.0122	0.0122	0.0177	0.0296

To us, it seems clear that the SAR specification with fixed effects is preferred. It addresses the spatial autocorrelation in the errors well, though not completely. This can be seen in the reduction in Moran's I statistic (even turning it negative) and in the Lagrange multiplier test for spatial lags in the dependent variable. We conjecture that with a less crude spatial weighting matrix, the test statistics could be improved.

The conclusion that should be drawn from our empirical analysis is that most of the spatial autocorrelation in the error is due to omission of the spatially lagged dependent variable. SAR does a good (though not perfect) job

[5] See Kapoor, Kelejian and Prucha (2007).

of correcting this problem. The SAR model specification yields a coefficient of 0.174 on the spatially lagged dependent variable, with a standard error of 0.014.

4 Implications

There are two implications of our analysis, the first obvious and the second more subtle.

The first implication is: *even if one wants to view the justification of the use of night lights data as one of a purely empirical proxy for GDP rather than a story about light demand or trade, OLS is not an appropriate specification due to spatial autocorrelation in the errors.* If one inserts basic demand and trade theory into the justification, then the case for misspecification is even stronger, as there is a theoretical argument for misspecification in addition to an empirical one.

The second implication is more subtle, as it is related to how the night lights data is used in the actual calculations to proxy for GDP data. In Henderson, Storeygard and Weil (2012), the regressions of growth in GDP on growth in night lights given in their Table 2 are not actually used for this purpose, but rather to justify using night lights as a proxy. In fact, what is used is a particular property of the OLS estimator that, when combined with sample moments, can be used to solve the parameters of the model, particularly β. That property, namely equation (4) of their paper, is reproduced here:[6]

$$\text{plim}\widehat{\psi} = \frac{1}{\beta}\left(\frac{\beta^2\sigma_y^2}{\beta^2\sigma_y^2 + \sigma_x^2}\right) \tag{4}$$

where σ_y^2 is the variance of true income growth and σ_x^2 is the variance of ε_x, defined in equation (2).[7] If our story here about model misspecification is correct, then the OLS estimator $\widehat{\psi}$ used in equation (4) should be replaced by a model incorporating spatial lags in the dependent variable and fixed effects. *Thus, the OLS estimator $\widehat{\psi}$ should be replaced by a SAR estimator with time and country fixed effects.* This has unknown implications for the calculations in

[6] For some reason, fixed effects and a constant term are ignored in this expression, though they are included in all the regressions.

[7] As stated on p. 1007 of Henderson, Storeygard and Weil (2012), equation (4) implies that $\widehat{\psi}$ is a biased estimate of the inverse elasticity of lights with respect to income. Our claim is that even after a correction is made using equation (4) and sample moments, the estimate is still biased.

subsequent parts of the paper that use equation (4), perhaps requiring further assumptions.

References

[1] Henderson, J.V., A. Storeygard and D.N. Weil (2012). "Measuring Economic Growth from Outer Space," *American Economic Review* 102, 994–1028.

[2] Kapoor, M., H.H. Kelejian and I. Prucha (2007). "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics* 140, 97-130.